UNITED STATES PATENT APPLICATION

PROGRAMMABLE MEMORY ADDRESS AND DECODE CIRCUITS WITH LOW TUNNEL BARRIER INTERPOLY INSULATORS

INVENTORS
Leonard Forbes

Schwegman, Lundberg, Woessner, & Kluth, P.A.
1600 TCF Tower
121 South Eighth Street
Minneapolis, Minnesota 55402
ATTORNEY DOCKET 1303.029US1
MICRON 01-0485

PROGRAMMABLE MEMORY ADDRESS AND DECODE CIRCUITS WITH LOW TUNNEL BARRIER INTERPOLY INSULATORS

Cross Reference To Related Applications

5   This application is related to the following co-pending, commonly assigned U.S. patent applications: "DRAM Cells with Repressed Memory Metal Oxide Tunnel Insulators," attorney docket no. 1303.019US1, serial number _____, "Programmable Array Logic or Memory Devices with Asymmetrical Tunnel Barriers," attorney docket no. 1303.020US1, serial number _____, "Dynamic

10   Electrically Alterable Programmable Memory with Insulating Metal Oxide Interpoly Insulators," attorney docket no. 1303.024US1, serial number _____, and "Flash Memory with Low Tunnel Barrier Interpoly Insulators," attorney docket no. 1303.014US1, serial number_____, "SRAM Cells with Repressed Floating Gate Memory, Metal Oxide Tunnel Interpoly Insulators," attorney docket no.

15   1303.028US1, serial number _____, "Flash Memory with Low Tunnel Barrier Interpoly Insulators," attorney docket no. 1303.014US1, serial number_____, which are filed on even date herewith and each of which disclosure is herein incorporated by reference.

20   ## Field of the Invention

The present invention relates generally to integrated circuits, and in particular to in-service programmable logic arrays with low tunnel barrier interpoly insulators.

25

Background of the Invention

Logic circuits are an integral part of digital systems, such as computers. Essentially, a logic circuit processes a number of inputs to produce a number of outputs for use by the digital system. The inputs and outputs are generally electronic signals that take on one of two "binary" values, a "high" logic value or a "low" logic value. The logic circuit manipulates the inputs using binary logic which describes, in a mathematical way, a given or desired relationship between the inputs and the outputs of the logic circuit.

Logic circuits that are tailored to the specific needs of a particular customer can be very expensive to fabricate on a commercial basis. Thus, general purpose very large scale integration (VLSI) circuits are defined. VLSI circuits serve as many logic roles as possible, which helps to consolidate desired logic functions. However, random logic circuits are still required to tie the various elements of a digital system together.

Several schemes are used to implement these random logic circuits. One solution is standard logic, such as transistor-transistor logic (TTL). TTL integrated circuits are versatile because they integrate only a relatively small number of commonly used logic functions. The drawback is that large numbers of TTL integrated circuits are typically required for a specific application. This increases the consumption of power and board space, and drives up the overall cost of the digital system.

One alternative to standard logic is fully custom logic integrated circuits. Custom logic circuits are precisely tailored to the needs of a specific application. This allows the implementation of specific circuit architectures that dramatically reduces the number of parts required for a system. However, custom logic devices require significantly greater engineering time and effort, which increases the cost to develop these circuits and may also delay the production of the end system.

A less expensive alternative to custom logic is the "programmable logic array." Programmable logic arrays take advantage of the fact that complex combinational logic functions can be reduced and simplified into various standard forms. For example, logical functions can be manipulated and reduced down to

5 traditional Sum of Products (SOP) form. In SOP form, a logical function uses just two types of logic functions that are implemented sequentially. This is referred to as two-level logic and can be implemented with various conventional logic functions, e.g., AND-OR, NAND-NAND, NOR-NOR.

One benefit of the programmable logic array is that it provides a regular,

10 systematic approach to the design of random, combinational logic circuits. A multitude of logical functions can be created from a common building block, e.g., an array of transistors. The logic array is customized or "programmed" by creating a specific metallization pattern to interconnect the various transistors in the array to implement the desired function.

15 Programmable logic arrays are fabricated using photolithographic techniques that allow semiconductor and other materials to be manipulated to form integrated circuits as is known in the art. These photolithographic techniques essentially use light that is focused through lenses and masks to define patterns in the materials with microscopic dimensions. The equipment and techniques that are used to

20 implement this photolithography provide a limit for the size of the circuits that can be formed with the materials. Essentially, at some point, the lithography cannot create a fine enough image with sufficient clarity to decrease the size of the elements of the circuit. In other words, there is a minimum dimension that can be achieved through conventional photolithography. This minimum dimension is

25 referred to as the "critical dimension" (CD) or minimum "feature size" (F) of the photolithographic process. The minimum feature size imposes one constraint on the size of the components of a programmable logic array. In order to keep up with the

demands for larger programmable logic arrays, designers search for ways to reduce the size of the components of the array.

As the density requirements become higher and higher in logic and memories it becomes more and more crucial to minimize device area. The programmable logic array (PLA) circuit in the NOR-NOR configuration is one example of an architecture for implementing logic circuits.

Flash memory cells are one possible solution for high density memory requirements. Flash memories include a single transistor, and with high densities would have the capability of replacing hard disk drive data storage in computer systems. This would result in delicate mechanical systems being replaced by rugged, small and durable solid-state memory packages, and constitute a significant advantage in computer systems. What is required then is a flash memory with the highest possible density or smallest possible cell area.

Flash memories have become widely accepted in a variety of applications ranging from personal computers, to digital cameras and wireless phones. Both INTEL and AMD have separately each produced about one billion integrated circuit chips in this technology.

The original EEPROM or EARPROM and flash memory devices described by Toshiba in 1984 used the interploy dielectric insulator for erase. (See generally, F. Masuoka et al., "A new flash EEPROM cell using triple polysilicon technology," IEEE Int. Electron Devices Meeting, San Francisco, pp. 464-67, 1984; F. Masuoka et al., "256K flash EEPROM using triple polysilicon technology," IEEE Solid-State Circuits Conf., Philadelphia, pp. 168-169, 1985). Various combinations of silicon oxide and silicon nitride were tried. (See generally, S. Mori et al., "reliable CVD inter-poly dialectics for advanced E&EEPROM," Symp. On VLSI Technology, Kobe, Japan, pp. 16-17, 1985). However, the rough top surface of the polysilicon floating gate resulted in, poor quality interpoly oxides, sharp points, localized high electric fields, premature breakdown and reliability problems.

Widespread use of flash memories did not occur until the introduction of the ETOX cell by INTEL in 1988. (See generally, US PATENT 4,780, 424, "Process for fabricating electrically alterable floating gate memory devices," 25 Oct. 1988; B. Dipert and L. Hebert, "Flash memory goes mainstream," IEEE Spectrum, pp. 48-51,

5   October, 1993; R. D. Pashley and S. K. Lai, "Flash memories, the best of two worlds," IEEE Spectrum, pp. 30-33, December 1989). This extremely simple cell and device structure resulted in high densities, high yield in production and low cost. This enabled the widespread use and application of flash memories anywhere a non-volatile memory function is required. However, in order to enable a

10   reasonable write speed the ETOX cell uses channel hot electron injection, the erase operation which can be slower is achieved by Fowler-Nordhiem tunneling from the floating gate to the source. The large barriers to electron tunneling or hot electron injection presented by the silicon oxide-silicon interface, 3.2 eV, result in slow write and erase speeds even at very high electric fields. The combination of very high

15   electric fields and damage by hot electron collisions in the oxide result in a number of operational problems like soft erase error, reliability problems of premature oxide breakdown and a limited number of cycles of write and erase.

Other approaches to resolve the above described problems include; the use of different floating gate materials, e.g. SiC, SiOC, GaN, and GaAlN, which exhibit a

20   lower work function (see Figure 1A), the use of structured surfaces which increase the localized electric fields (see Figure 1B), and amorphous SiC gate insulators with larger electron affinity, χ, to increase the tunneling probability and reduce erase time (see Figure 1C).

One example of the use of different floating gate (Figure 1A) materials is

25   provided in US Patent no. 5,801,401 by L. Forbes, entitled "FLASH MEMORY WITH MICROCRYSTALLINE SILICON CARBIDE AS THE FLOATING GATE STRUCTURE." Another example is provided in US Patent no. 5,852,306 by L. Forbes, entitled "FLASH MEMORY WITH NANOCRYSTALLINE SILICON

FILM AS THE FLOATING GATE." Still further examples of this approach are provided in pending applications by L. Forbes and K. Ahn, entitled "DYNAMIC RANDOM ACCESS MEMORY OPERATION OF A FLASH MEMORY DEVICE WITH CHARGE STORAGE ON A LOW ELECTRON AFFINITY GaN OR GaAlN FLOATING GATE," serial no. 08/908098, and "VARIABLE ELECTRON AFFINITY DIAMOND-LIKE COMPOUNDS FOR GATES IN SILICON CMOS MEMORIES AND IMAGING DEVICES," serial no. 08/903452.

An example of the use of the structured surface approach (Figure 1B) is provided in US Patent no. 5,981,350 by J. Geusic, L. Forbes, and K.Y. Ahn, entitled "DRAM CELLS WITH A STRUCTURE SURFACE USING A SELF STRUCTURED MASK." Another example is provided in US Patent no. 6,025, 627 by L. Forbes and J. Geusic, entitled "ATOMIC LAYER EXPITAXY GATE INSULATORS AND TEXTURED SURFACES FOR LOW VOLTAGE FLASH MEMORIES."

Finally, an example of the use of amorphous SiC gate insulators (Figure 1C) is provided in US Patent Application serial no. 08/903453 by L. Forbes and K. Ahn, entitled "GATE INSULATOR FOR SILICON INTEGRATED CIRCUIT TECHNOLOGY BY THE CARBURIZATION OF SILICON."

Additionally, graded composition insulators to increase the tunneling probability and reduce erase time have been described by the same inventors. (See, L. Forbes and J. M. Eldridge, "GRADED COMPOSITION GATE INSULATORS TO REDUCE TUNNELING BARRIERS IN FLASH MEMORY DEVICES," application serial no. _____.

However, all of these approaches relate to increasing tunneling between the floating gate and the substrate such as is employed in a conventional ETOX device and do not involve tunneling between the control gate and floating gate through and inter-poly dielectric.

Therefore, there is a need in the art to provide improved programmable memory address and decode circuits. The programmable memory address and decode circuits should provide improved flash memory densities while avoiding the large barriers to electron tunneling or hot electron injection presented by the silicon oxide-silicon interface, 3.2 eV, which result in slow write and erase speeds even at very high electric fields. There is also a need to avoid the combination of very high electric fields and damage by hot electron collisions in the which oxide result in a number of operational problems like soft erase error, reliability problems of premature oxide breakdown and a limited number of cycles of write and erase. Further, when using an interpoly dielectric insulator erase approach, the above mentioned problems of having a rough top surface on the polysilicon floating gate which results in, poor quality interpoly oxides, sharp points, localized high electric fields, premature breakdown and reliability problems must be avoided.

Summary of the Invention

The above mentioned problems with programmable memory address and decode circuits and other problems are addressed by the present invention and will be understood by reading and studying the following specification. Systems and methods are provided for programmable memory address and decode circuits using logic cells, or non-volatile memory cells with metal oxide and/or low tunnel barrier interpoly insulators.

In one embodiment of the present invention, programmable memory address and decode circuits with low tunnel tunnel barrier interpoly insulators having a tunnel barrier of approximately 2.0 eV are provided. The decoder for a memory device includes a number of address lines and a number of output lines wherein the address lines and the output lines form an array. A number of logic cells are formed at the intersections of output lines and address lines. Each of the logic cells includes a floating gate transistor which includes a first source/drain region and a second

source/drain region separated by a channel region in a substrate. A floating gate opposes the channel region and is separated therefrom by a gate oxide. A control gate opposing the floating gate. The control gate is separated from the floating gate by the low tunnel barrier intergate insulator. The low tunnel barrier intergate

5   insulator includes a metal oxide insulator selected from the group consisting of PbO, Al_2O_3, Ta_2O_5, TiO_2, ZrO_2, Nb_2O_5 and/or a Perovskite oxide tunnel barrier. The floating gate includes a polysilicon floating gate having a metal layer formed thereon in contact with the low tunnel barrier intergate insulator. And, the control gate includes a polysilicon control gate having a metal layer formed thereon in

10   contact with the low tunnel barrier intergate insulator.

These and other embodiments, aspects, advantages, and features of the present invention will be set forth in part in the description which follows, and in part will become apparent to those skilled in the art by reference to the following description of the invention and referenced drawings or by practice of the invention.

15   The aspects, advantages, and features of the invention are realized and attained by means of the instrumentalities, procedures, and combinations particularly pointed out in the appended claims.

Brief Description of the Drawings

20   Figures 1A-1C illustrate a number of previous methods for reducing tunneling barriers in Flash memory.

Figure 2 illustrates one embodiment of a floating gate transistor, or non-volatile memory cell, according to the teachings of the present invention.

Figure 3 illustrates another embodiment of a floating gate transistor, or non-

25   volatile memory cell, according to the teachings of the present invention.

Figure 4 is a perspective view illustrating an array of silicon pillars formed on a substrate as used in one embodiment according to the teachings of the present invention.

Figures 5A-5E are cross sectional views taken along cut line 5-5 from Figure 4 illustrating a number of floating gate and control gate configurations which are included in the scope of the present invention.

Figures 6A-6C illustrate a number of address coincidence schemes can be used together with the present invention.

Figure 7A is an energy band diagram illustrating the band structure at vacuum level with the low tunnel barrier interpoly insulator according to the teachings of the present invention.

Figure 7B is an energy band diagram illustrating the band structure during an erase operation of electrons from the floating gate to the control gate across the low tunnel barrier interpoly insulator according to the teachings of the present invention.

Figure 7C is a graph plotting tunneling currents versus the applied electric fields (reciprocal applied electric field shown) for an number of barrier heights.

Figure 8 shows a conventional NOR decode array for memory circuits according to the teachings of the prior art.

Figure 9 is a schematic diagram illustrating an embodiment of a decode circuit, or memory address decoder, according to the teachings of the present invention.

Figure 10 is a simplified block diagram of a high-level organization of an electronic system according to the teachings of the present invention.

Description of the Preferred Embodiments

In the following detailed description of the invention, reference is made to the accompanying drawings which form a part hereof, and in which is shown, by way of illustration, specific embodiments in which the invention may be practiced. The embodiments are intended to describe aspects of the invention in sufficient detail to enable those skilled in the art to practice the invention. Other embodiments may be utilized and changes may be made without departing from the scope of the

present invention. In the following description, the terms wafer and substrate are interchangeably used to refer generally to any structure on which integrated circuits are formed, and also to such structures during various stages of integrated circuit fabrication. Both terms include doped and undoped semiconductors, epitaxial layers

5   of a semiconductor on a supporting semiconductor or insulating material, combinations of such layers, as well as other such structures that are known in the art. The following detailed description is not to be taken in a limiting sense, and the scope of the present invention is defined only by the appended claims.

The term "horizontal" as used in this application is defined as a plane

10   parallel to the conventional plane or surface of a wafer or substrate, regardless of the orientation of the wafer or substrate. The term "vertical" refers to a direction perpendicular to the horizonal as defined above. Prepositions, such as "on", "side" (as in "sidewall"), "higher", "lower", "over" and "under" are defined with respect to the conventional plane or surface being on the top surface of the wafer or substrate,

15   regardless of the orientation of the wafer or substrate. The following detailed description is, therefore, not to be taken in a limiting sense, and the scope of the present invention is defined only by the appended claims, along with the full scope of equivalents to which such claims are entitled.

The present invention, describes the use of metal oxide inter-poly dielectric

20   insulators between the control gate and the floating gate. An example is shown in Figure 2 for a planar structure, or horizontal non-volatile memory cell. According to the teachings of the present invention. The use of metal oxide films for this purpose offer a number of advantages including:

(i)   Flexibility in selecting a range of smooth metal film surfaces and

25   compositions that can be oxidized to form tunnel barrier insulators.

(ii)   Employing simple "low temperature oxidation" to produce oxide films of highly controlled thickness, composition, purity and uniformity.

(iii)   Avoiding inadvertent inter-diffusion of the metal and silicon as well as silicide formation since the oxidation can be carried out at such low temperatures.

(iv)   Using metal oxides that provide desirably lower tunnel barriers, relative to barriers currently used such as SiO_2.

(v)   Providing a wide range of higher dielectric constant oxide films with improved capacitance characteristics.

(vi)   Providing a unique ability to precisely tailor tunnel oxide barrier properties for various device designs and applications.

(vii)   Permitting the use of thicker tunnel barriers, if needed, to enhance device performance and its control along with yield and reliability.

(viii)   Developing layered oxide tunnel barriers by oxidizing layered metal film compositions in order, for example, to enhance device yields and reliability more typical of single insulating layers.

(ix)   Eliminating soft erase errors caused by the current technique of tunnel erase from floating gate to the source.

Figure 2 illustrates one embodiment of a floating gate transistor, or non-volatile memory cell 200, according to the teachings of the present invention. As shown in Figure 2, the non-volatile memory cell 200 includes a first source/drain region 201 and a second source/drain region 203 separated by a channel region 205 in a substrate 206. A floating gate 209 opposes the channel region 205 and is separated therefrom by a gate oxide 211. A control gate 213 opposes the floating gate 209. According to the teachings of the present invention, the control gate 213 is separated from the floating gate 209 by a low tunnel barrier intergate insulator 215.

In one embodiment of the present invention, low tunnel barrier intergate insulator 215 includes a metal oxide insulator selected from the group consisting of lead oxide (PbO) and aluminum oxide (Al_2O_3). In an alternative embodiment of the present invention, the low tunnel barrier intergate insulator 215 includes a transition

metal oxide and the transition metal oxide is selected from the group consisting of Ta_2O_5, TiO_2, ZrO_2, and Nb_2O_5. In still another alternative embodiment of the present invention, the low tunnel barrier intergate insulator 215 includes a Perovskite oxide tunnel barrier.

5   According to the teachings of the present invention, the floating gate 209 includes a polysilicon floating gate 209 having a metal layer 216 formed thereon in contact with the low tunnel barrier intergate insulator 215. Likewise, the control gate 213 includes a polysilicon control gate 213 having a metal layer 217 formed thereon in contact with the low tunnel barrier intergate insulator 215. In this

10   invention, the metal layers, 216 and 217, are formed of the same metal material used to form the metal oxide interpoly insulator 215.

Figure 3 illustrates another embodiment of a floating gate transistor, or non-volatile memory cell 300, according to the teachings of the present invention. As shown in the embodiment of Figure 3, the non-volatile memory cell 300 includes a

15   vertical non volatile memory cell 300. In this embodiment, the non-volatile memory cell 300 has a first source/drain region 301 formed on a substrate 306. A body region 307 including a channel region 305 is formed on the first source/drain region 301. A second source/drain region 303 is formed on the body region 307. Methods for forming such a vertical transistor structure are disclosed in US Patent

20   no. 6,135,175, entitled "Memory Address Decode Array with vertical transistors, which is incorporated herein by reference. A floating gate 309 opposes the channel region 305 and is separated therefrom by a gate oxide 311. A control gate 313 opposes the floating gate 309. According to the teachings of the present invention, the control gate 313 is separated from the floating gate 309 by a low tunnel barrier

25   intergate insulator 315.

According to the teachings of the present invention, the low tunnel barrier intergate insulator 315 includes a metal oxide insulator 315 selected from the group consisting of PbO, Al_2O_3, Ta_2O_5, TiO_2, ZrO_2, and Nb_2O_5. In still another alternative

embodiment of the present invention, the low tunnel barrier intergate insulator 315 includes a Perovskite oxide tunnel barrier. The floating gate 309 includes a polysilicon floating gate 309 having a metal layer 316 formed thereon in contact with the low tunnel barrier intergate insulator 315. The control gate 313 includes a polysilicon control gate 313 having a metal layer 317 formed thereon in contact with the low tunnel barrier intergate insulator 315.

As shown in Figure 3, the floating gate 309 includes a vertical floating gate 309 formed alongside of the body region 307. In the embodiment shown in Figure 3, the control gate 313 includes a vertical control gate 313 formed alongside of the vertical floating gate 309.

As will be explained in more detail below, the floating gate 309 and control gate 313 orientation shown in Figure 3 is just one embodiment for a vertical non volatile memory cell 300, according to the teachings of the present invention. In other embodiments, explained below, the floating gate includes a horizontally oriented floating gate formed alongside of the body region. In this alternative embodiment, the control gate includes a horizontally oriented control gate formed above the horizontally oriented floating gate.

Figure 4 is a perspective view illustrating an array of silicon pillars 400-1, 400-2, 400-3, . . ., 400-N, formed on a substrate 406 as used in one embodiment according to the teachings of the present invention. As will be understood by one of ordinary skill in the art upon reading this disclosure, the substrates can be (i) conventional p-type bulk silicon or p-type epitaxial layers on p+ wafers, (ii) silicon on insulator formed by conventional SIMOX, wafer bonding and etch back or silicon on sapphire, or (iii) small islands of silicon on insulator utilizing techniques such as described in more detail in U.S. patent no. 5,691,230, by Leonard Forbes, entitled "Technique for Producing Small Islands of Silicon on Insulator," issued 11/25/1997, which is incorporated herein by reference.

As shown in Figure 4, each pillar in the array of silicon pillars 400-1, 400-2, 400-3, . . ., 400-N, includes a first source/drain region 401 and a second source/drain region 403. The first and the second source/drain regions, 401 and 403, are separated by a body region 407 including channel regions 405. As shown in Figure 4, a number of trenches 430 separate adjacent pillars in the array of silicon pillars 400-1, 400-2, 400-3, . . ., 400-N. Trenches 430 are referenced in connection with the discussion which follows in connection with Figures 5A-5E.

Figures 5A-5E are cross sectional views taken along cut line 5-5 from Figure 4. As mentioned above in connection with Figure 3, a number of floating gate and control gate configurations are included in the present invention. Figure 5A illustrates one such embodiment of the present invention. Figure 5A illustrates a first source/drain region 501 and second source/drain region 503 for a non-volatile memory cell 500 formed according to the teachings of the present invention. As shown in Figure 5, the first and second source/drain regions, 501 and 503, are contained in a pillar of semiconductor material, and separated by a body region 507 including channel regions 505. As shown in the embodiments of Figures 5A-5E, the first source/drain region 501 is integrally connected to a buried sourceline 525. As one or ordinary skill in the art will understand upon reading this disclosure the buried sourceline 525 is be formed of semiconductor material which has the same doping type as the first source/drain region 501. In one embodiment, the sourceline 525 is formed of semiconductor material of the same doping as the first source/drain region 501, but is more heavily doped than the first source/drain region 501.

As shown in the embodiment of Figure 5A, a pair of floating gates 509-1 and 509-2 are formed in each trench 530 between adjacent pillars which form memory cells 500-1 and 500-2. Each one of the pair of floating gates, 509-1 and 509-2, respectively opposes the body regions 507-1 and 507-2 in adjacent pillars 500-1 and 500-2 on opposing sides of the trench 530.

-14-

In this embodiment, a single control gate 513 is shared by the pair of floating gates 509-1 and 509-2 on opposing sides of the trench 530. As one of ordinary skill in the art will understand upon reading this disclosure, the shared single control gate 513 can include an integrally formed control gate line. As shown in Figure 5A, such an integrally formed control gate line 513 can be one of a plurality of control gate lines which are each independently formed in the trench, such as trench 530, below the top surface of the pillars 500-1 and 500-2 and between the pair of floating gates 509-1 and 509-2. In one embodiment, according to the teachings of the present invention, each floating gate, e.g. 509-1 and 509-2, includes a vertically oriented floating gate having a vertical length of less than 100 nanometers.

As shown in the embodiment of Figure 5B, a pair of floating gates 509-1 and 509-2 are formed in each trench 530 between adjacent pillars which form memory cells 500-1 and 500-2. Each one of the pair of floating gates, 509-1 and 509-2, respectively opposes the body regions 507-1 and 507-2 in adjacent pillars 500-1 and 500-2 on opposing sides of the trench 530.

In the embodiment of Figure 5B, a plurality of control gate lines are again formed in trenches, e.g. trench 530, below the top surface of the pillars, 500-1 and 500-2, and between the pair of floating gates 509-1 and 509-2. However, in this embodiment, each trench, e.g. 530, houses a pair of control gate lines, shown as 513-1 and 513-2. Each one of the pair of control gate lines 513-1 and 513-2 adresses the floating gates, 509-1 and 509-2 respectively, on opposing sides of the trench 530. In this embodiment, the pair of control gate lines, or control gates 513-1 and 513-2 are separated by an insulator layer.

As shown in the embodiment of Figure 5C, a pair of floating gates 509-1 and 509-2 are again formed in each trench 530 between adjacent pillars which form memory cells 500-1 and 500-2. Each one of the pair of floating gates, 509-1 and 509-2, respectively opposes the body regions 507-1 and 507-2 in adjacent pillars 500-1 and 500-2 on opposing sides of the trench 530.

In the embodiment of Figure 5C, the plurality of control gate lines are disposed vertically above the floating gates. That is, in one embodiment, the control gate lines are located above the pair of floating gates 509-1 and 509-2 and not fully beneath the top surface of the pillars 500-1 and 500-2. In the embodiment of Figure 5C, each pair of floating gates, e.g. 509-1 and 509-2, in a given trench shares a single control gate line, or control gate 513.

As shown in the embodiment of Figure 5D, a pair of floating gates 509-1 and 509-2 are formed in each trench 530 between adjacent pillars which form memory cells 500-1 and 500-2. Each one of the pair of floating gates, 509-1 and 509-2, respectively opposes the body regions 507-1 and 507-2 in adjacent pillars 500-1 and 500-2 on opposing sides of the trench 530.

In the embodiment of Figure 5D, the plurality of control gate lines are disposed vertically above the floating gates. That is, in one embodiment, the control gate lines are located above the pair of floating gates 509-1 and 509-2 and not fully beneath the top surface of the pillars 500-1 and 500-2. However, in the embodiment of Figure 5D, each one of the pair of floating gates, e.g. 509-1 and 509-2, is addressed by an independent one of the plurality of control lines or control gates, shown in Figure 5D as 513-1 and 513-2.

As shown in the embodiment of Figure 5E, a single floating gate 509 is formed in each trench 530 between adjacent pillars which form memory cells 500-1 and 500-2. According to the teachings of the present invention, the single floating gate 509 can be either a vertically oriented floating gate 509 or a horizontally oriented floating gate 509 formed by conventional processing techniques, or can be a horizontally oriented floating gate 509 formed by a replacement gate technique such as described in a copending application, entitled "Flash Memory with Ultrathin Vertical Body Transistors," by Leonard Forbes and Kie Y. Ahn, application serial no. 09/780,169. In one embodiment of the present invention, the floating gate 509 has a vertical length facing the body region 505 of less than 100 nm. In another

embodiment, the floating gate 509 has a vertical length facing the body region 505 of less than 50 nm. In one embodiment, as shown in Figure 5E, the floating gate 509 is shared, respectively, with the body regions 507-1 and 507-2, including channel regions 505-1 and 505-2, in adjacent pillars 500-1 and 500-2 located on opposing sides of the trench 530. And, as shown in Figure 5E, the control gate includes a single horizontally oriented control gate line, or control gate 513 formed above the horizontally oriented floating gate 509.

As one of ordinary skill in the art will understand upon reading this disclosure, in each of the embodiments described above in connection with Figures 5A-5E the floating gates 509 are separated from the control gate lines, or control gates 513 with a low tunnel barrier intergate insulator in accordance with the descriptions given above in connection with Figure 3. The modifications here are to use tunneling through the interpoly dielectric to realize flash memory devices. The vertical devices include an extra flexibility in that the capacitors, e.g. gate oxide and intergate insulator, are easily fabricated with different areas. This readily allows the use of very high dielectric constant inter-poly dielectric insulators with lower tunneling barriers.

Figures 6A-6C illustrate that a number of address coincidence schemes can be used together with the present invention. Figure 6A illustrates a NOR flash memory array 610 having a number of non-volatile memory cells 600-1, 600-2, 600-3, using a coincidence address array scheme. For purposes of illustration, Figure 6A shows a sourceline 625 coupled to a first source/drain region 601 in each of the number of non-volatile memory cells 600-1, 600-2, 600-3. The sourceline is shown oriented in a first selected direction in the flash memory array 610. In Figure 6A, a number of control gate lines 630 are shown oriented in a second selected direction in the flash memory array 610. As shown in Figure 6A, the number of control gate lines 630 are coupled to, or integrally formed with the control gates 613 for the number of non-volatile memory cells 600-1, 600-2, 600-3. As shown in Figure 6A,

-17-

the second selected direction is orthogonal to the first selected direction. Finally, Figure 6A shows a number of bitlines 635 oriented in a third selected direction in the flash memory array 610. As shown in Figure 6A, the number of bitlines are coupled to the second source/drain regions 603 in the number of non-volatile

5   memory cells 600-1, 600-2, 600-3. In the embodiment shown in Figure 6A the third selected direction is parallel to the second selected direction and the number of control gate lines 630 serve as address lines. Also, as shown in Figure 6A, the flash memory array 610 includes a number of backgate or substrate/well bias address lines 640 coupled to the substrate.

10   Using Figure 6A as a reference point, Figures 6B-6C illustrate of top view for three different coincidence address scheme layouts suitable for use with the present invention. First, Figure 6B provides the top view layout of the coincidence address scheme described in connection with Figure 6A. This is, Figure 6B illustrates a number of sourcelines 625 oriented in a first selected direction, a

15   number of control gate lines 630 oriented in a second selected direction, and a number of bitlines 635 oriented in a third selected direction for the flash memory array 600. In the embodiment of Figure 6B, the first selected direction and the third selected direction are parallel to one another and orthogonal to the second selected direction. In this embodiment, the number of control gate lines 630 serve as address

20   lines. According to the teachings of the present invention, the output lines, e.g. bitlines 635 are perpendicular to the address lines, e.g. in this embodiment control gate lines 630.

Figure 6C provides the top view layout of yet another coincidence address scheme according to the teachings of the present invention. This is, Figure 6C

25   illustrates a number of sourcelines 625 oriented in a first selected direction, a number of control gate lines 630 oriented in a second selected direction, and a number of bitlines 635 oriented in a third selected direction for the flash memory array 600. In the embodiment of Figure 6C, the first selected direction and the

second selected direction are parallel to one another and orthogonal to the third

selected direction. In this embodiment, the number of bitlines 635 serve as output

data lines. In an alternative embodiment, the sourcelines 625 can include a uniform

ground plane as the same will be known and understood by one of ordinary skill in

5   the art.

As will be apparent to one of ordinary skill in the art upon reading this

disclosure, and as will be described in more detail below, write can still be achieved

by hot electron injection and/or, according to the teachings of the present invention,

tunneling from the control gate. According to the teachings of the present invention,

10   block erase is accomplished by driving the control gates with a relatively large

positive voltage and tunneling from the metal on top of the floating gate to the metal

on the bottom of the control gate.

Figure 7A is an energy band diagram illustrating the band structure at

vacuum level with the low tunnel barrier interpoly insulator according to the

15   teachings of the present invention. Figure 7A is useful in illustrating the reduced

tunnel barrier off of the floating gate to the control gate and for illustrating the

respective capacitances of the structure according to the teachings of the present

invention.

Figure 7A shows the band structure of the silicon substrate, e.g. channel

20   region 701, silicon dioxide gate insulator, e.g. gate oxide 703, polysilicon floating

gate 705, the low tunnel barrier interpoly dielectric 707, between metal plates 709

and 711, and then the polysilicon control gate 713, according to the teachings of the

present invention.

The design considerations involved are determined by the dielectric constant,

25   thickness and tunneling barrier height of the interpoly dielectric insulator 707

relative to that of the silicon dioxide gate insulator, e.g. gate oxide 703. The

tunneling probability through the interpoly dielectric 707 is an exponential function

of both the barrier height and the electric field across this dielectric.

Figure 7B is an energy band diagram illustrating the band structure during an erase operation of electrons from the floating gate 705 to the control gate 713 across the low tunnel barrier interpoly insulator 707 according to the teachings of the present invention. Figure 7B is similarly useful in illustrating the reduced tunnel barrier off of the floating gate to the control gate and for illustrating the respective capacitances of the structure according to the teachings of the present invention.

As shown in Figure 7B, the electric field is determined by the total voltage difference across the structure, the ratio of the capacitances (see Figure 7A), and the thickness of the interpoly dielectric 707. The voltage across the interpoly dielectric 707 will be, $\Delta V2 = V\ C1/(C1 + C2)$, where V is the total applied voltage. The capacitances, C, of the structures depends on the dielectric constant, ϵ_r, the permittivity of free space, ϵ_o, and the thickness of the insulating layers, t, and area, A, such that $C = \epsilon_r\ \epsilon_o\ A/t$, Farads/cm^2, where ϵ_r represents the low frequency dielectric constant.. The electric field across the interpoly dielectric insulator 707, having capacitance, C2, will then be $E2 = \Delta V2/t2$, where t2 is the thickness of this layer.

The tunneling current in erasing charge from the floating gate 705 by tunneling to the control gate 713 will then be as shown in Figure 7B given by an equation of the form:

$$J = B\ \exp(-\ Eo/E)$$

where E is the electric field across the interpoly dielectric insulator 707 and Eo depends on the barrier height. Practical values of current densities for aluminum oxide which has a current density of 1 A/cm^2 at a field of about E = 1V/20A = $5\times10^{+6}$ V/cm are evidenced in a description by Pollack. (See generally, S. R. Pollack and C. E. Morris, "Tunneling through gaseous oxidized films of Al_2O_3," Trans. AIME, Vol. 233, p. 497, 1965). Practical current densities for silicon oxide transistor gate insulators which has a current density of 1 A/cm^2 at a field of about E = 2.3V/23A = $1\times10^{+7}$ V/cm are evidenced in a description by T. P. Ma et al. (See

generally, T. P. Ma et al., "Tunneling leakage current in ultrathin (<a4 nm)
nitride/oxide stack dielectrics," IEEE Electron Device Letters, vol. 19, no. 10, pp.
388-390, 1998).

The lower electric field in the aluminum oxide interpoly insulator 707 for the
same current density reflects the lower tunneling barrier of less than 2 eV, shown in
Figure 7B, as opposed to the 3.2 eV tunneling barrier of silicon oxide 703, also
illustrated in Figure 7B.

Figure 7C is a graph plotting tunneling currents versus the applied electric
fields (reciprocal applied electric field shown) for an number of barrier heights.
Figure 7C illustrates the dependence of the tunneling currents on electric field
(reciprocal applied electric field) and barrier height. The fraction of voltage across
the interpoly or intergate insulator, $\Delta V2$, can be increased by making the area of the
intergate capacitor, C2, (e.g. intergate insulator 707) smaller than the area of the
transistor gate capacitor, C1 (e.g. gate oxide 703). This would be required with high
dielectric constant intergate dielectric insulators 707 and is easily realized with the
vertical floating gate structures described above in connection with Figures 3, and
5A-5E.

Methods of Formation

Several examples are outlined below in order to illustrate how a diversity of
such metal oxide tunnel barriers can be formed, according to the teachings of the
present invention. Processing details and precise pathways taken which are not
expressly set forth below will be obvious to one of ordinary skill in the art upon
reading this disclosure. Firstly, although not included in the details below, it is
important also to take into account the following processing factors in connection
with the present invention:

(i)   The poly-Si layer is to be formed with emphasis on obtaining a
surface that is very smooth and morphologically stable at subsequent device
processing temperatures which will exceed that used to grow Metal oxide.

(ii)   The native SiO_x oxide on the poly-Si surface must be removed (e.g.,
by sputter cleaning in an inert gas plasma *in situ*) just prior to depositing the metal
film. The electrical characteristics of the resultant Poly-Si/Metal/Metal
oxide/Metal/Poly-Si structure will be better defined and reproducible than that of a
Poly-Si/Native SiO_x/Metal/Metal oxide/Poly-Si structure.

(iii)   The oxide growth rate and limiting thickness will increase with
oxidation temperature and oxygen pressure. The oxidation kinetics of a metal may,
in some cases, depend on the crystallographic orientations of the very small grains
of metal which comprise the metal film (see generally, O, Kubaschewski and B. E.
Hopkins, "Oxidation of Metals and Alloys", Butterworth, London, pp. 53-64, 1962).
If such effects are significant, the metal deposition process can be modified in order
to increase its preferred orientation and subsequent oxide thickness and tunneling
uniformity. To this end, use can be made of the fact that metal films strongly prefer
to grow during their depositions having their lowest free energy planes parallel to
the film surface. This preference varies with the crystal structure of the metal. For
example, fcc metals prefer to form {111} surface plans. Metal orientation effects, if
present, would be larger when only a limited fraction of the metal will be oxidized
and unimportant when all or most of the metal is oxidized.

(iv)   Modifications in the structure shown in Figure 2 may be introduced
in order to compensate for certain properties in some metal/oxide/metal layers.
Such changes are reasonable since a wide range of metals, alloys and oxides with
quite different physical and chemical properties can be used to form these tunnel
junctions.

Example I. Formation of PbO Tunnel Barriers

This oxide barrier has been studied in detail using Pb/PbO/Pb structures. The oxide itself can be grown very controllably on deposited lead films using either thermal oxidation (see generally, J. M. Eldridge and J. Matisoo, "Measurement of

5   tunnel current density in a Meal-Oxide-Metal system as a function of oxide thickness," Proc. 12th Intern. Conf. on Low Temperature Physics, pp. 427-428, 1971; J. M. Eldridge and D. W. Dong, "Growth of thin PbO layers on lead films. I. Experiment," Surface Science, Vol. 40, pp. 512-530, 1973) or rf sputter etching in an oxygen plasma (see generally, J. H. Greiner, "Oxidation of lead films by rf

10   sputter etching in an oxygen plasma", J. Appl. Phys., Vol. 45, No. 1, pp. 32-37, 1974). It will be seen that there are a number of possible variations on this structure. Starting with a clean poly-Si substrate, one processing sequence using thermal oxidation involves:

(i)   Depositing a clean lead film on the poly-Si floating gate at ~25 to

15   75C in a clean vacuum system having a base pressure of ~10^{-8} Torr or lower. The Pb film will be very thin with a thickness within 1 or 2A of its target value.

(ii)   Lead and other metal films can be deposited by various means including physical sputtering and/or from a Knudsen evaporation cell. The sputtering process also offers the ability to produce smoother films by increasing the

20   re-sputtering-to-deposition ratio since re-sputtering preferentially reduces geometric high points of the film.

(iii)   Using a "low temperature oxidation process" to grow an oxide film of self-limited thickness. In this case, oxygen gas is introduced at the desired pressure in order to oxidize the lead *in situ* without an intervening exposure to

25   ambient air. For a fixed oxygen pressure and temperature, the PbO thickness increases with log(time). Its thickness can be controlled via time or other parameters to *within 0.10 A*, as determined via *in situ* ellipsometric or *ex situ* measurements of Josephson tunneling currents. This control is demonstrated by the

-23-

very limited statistical scatter of the current PbO thickness data shown in the insert

of Fig. 3 in an article by J. M. Eldridge and J. Matisoo, entitled "Measurement of

tunnel current density in a Meal-Oxide-Metal system as a function of oxide

thickness," Proc. 12th Intern. Conf. on Low Temperature Physics, pp. 427-428, 1971.

5   This remarkable degree of control over tunnel current is due to the excellent control

over PbO thickness that can be achieved by "low temperature oxidation." For

example, increasing the oxidation time from 100 to 1,000 minutes at an oxygen

pressure of 750 Torr at 25C only raises the PbO thickness by 3 A (e.g., from ~21 to

24 A, see Fig. 1 in J. M. Eldridge and J. Matisoo, "Measurement of tunnel current

10   density in a Meal-Oxide-Metal system as a function of oxide thickness," Proc. 12th

Intern. Conf. on Low Temperature Physics, pp. 427-428, 1971). Accordingly,

controlling the oxidation time to within 1 out of a nominal 100 minute total

oxidation time provides a thickness that is within 0.1 A of 21A. The PbO has a

highly stoichiometric composition throughout its thickness, as evidenced from

15   ellipsometry (e.g., see Fig. 6 in J. M. Eldridge and D. W. Dong, "Growth of thin

PbO layers on lead films. I. Experiment," Surface Science, Vol. 40, pp. 512-530,

1973) and the fact that the tunnel barrier heights are identical for Pb/PbO/Pb

structures.

(iv)   Re-evacuate the system and deposit the top lead electrode. This

20   produces a tunnel structure having virtually identical tunnel barriers at both Pb/O

interfaces.

(v)   The temperature used to subsequently deposit the Poly-Si control

gate must be held below the melting temperature (327C) of lead. The PbO itself is

stable (up to ~500C or higher) and thus introduces no temperature constraint on

25   subsequent processes. One may optionally oxidize the lead film to completion,

thereby circumventing the low melting temperature of metallic lead. In this case,

one would form a Poly-Si/PbO/Poly-Si tunnel structure having an altered tunnel

barrier for charge injection. Yet another variation out of several would involve:

oxidizing the lead film to completion; replacing the top lead electrode with a higher

melting metal such as Al; and, then adding the poly-Si control layer. This junction

would have asymmetrical tunneling behavior due to the difference in barrier heights

between the Pb/PbO and PbO/Al electrodes.

5

Example II. Formation of Al_2O_3 Tunnel Barriers

A number of studies have dealt with electron tunneling in $Al/Al_2O_3/Al$

structures where the oxide was grown by "low temperature oxidation" in either

molecular or plasma oxygen (see generally, S. M. Sze, Physics of Semiconductor

10 Devices, Wiley, NY, pp. 553-556, 1981; G. Simmons and A. El-Badry,

"Generalized formula for the electric tunnel effect between similar electrodes

separated by a thin insulating film," J. Appl. Phys., Vol. 34, p. 1793, 1963; S. R.

Pollack and C. E. Morris, "Tunneling through gaseous oxidized films of Al_2O_3,"

Trans. AIME, Vol. 233, p. 497, 1965; Z. Hurych, "Influence of nonuniform

15 thickness of dielectric layers on capacitance and tunnel currents," Solid-State

Electronics, Vol. 9, p. 967, 1966; S. P. S. Arya and H. P. Singh, "Conduction

properties of thin Al_2O_3 films," Thin Solid Films, Vol. 91, No. 4, pp. 363-374, May

1982; K.-H. Gundlach and J. Holzl, "Logarithmic conductivity of $Al-Al_2O_3-Al$

tunneling junctions produced by plasma- and by thermal-oxidation", surface

20 Science, Vol. 27, pp. 125-141, 1971). Before sketching out a processing sequence

for these tunnel barriers, note:

(i)   Capacitance and tunnel measurements indicate that the Al_2O_3

thickness increases with the log (oxidation time), similar to that found for PbO/Pb as

well as a great many other oxide/metal systems.

25   (ii)   Tunnel currents are asymmetrical in this system with somewhat

larger currents flowing when electrons are injected from Al/Al_2O_3 interface

developed during oxide growth. This asymmetry is due to a minor change in

composition of the growing oxide: there is a small concentration of excess metal in

the Al_2O_3, the concentration of which diminishes as the oxide is grown thicker. The excess Al^{+3} ions produce a space charge that lowers the tunnel barrier at the inner interface. The oxide composition at the outer Al_2O_3/Al contact is much more stoichiometric and thus has a higher tunnel barrier. *In situ* ellipsometer

5 measurements on the thermal oxidation of Al films deposited and oxidized *in situ* support this model (see generally, J. Grimblot and J. M. Eldridge, "I. Interaction of Al films with O_2 at low pressures", J. Electro. Chem. Soc., Vol. 129, No. 10, pp. 2366-2368, 1982. J. Grimblot and J. M. Eldridge, "II. Oxidation of Al films", ibid, 2369-2372, 1982). In spite of this minor complication, $Al/Al_2O_3/Al$ tunnel barriers

10 can be formed that will produce predictable and highly controllable tunnel currents that can be ejected from either electrode. The magnitude of the currents are still primarily dominated by Al_2O_3 thickness which can be controlled via the oxidation parametrics.

With this background, we can proceed to outline one process path out of

15 several that can be used to form Al_2O_3 tunnel barriers. Here the aluminum is thermally oxidized although one could use other techniques such as plasma oxidation (see generally, S. R. Pollack and C. E. Morris, "Tunneling through gaseous oxidized films of Al_2O_3," Trans. AIME, Vol. 233, p. 497, 1965; K. -H. Gundlach and J. Holzl, "Logarithmic conductivity of $Al-Al_2O_3-Al$ tunneling

20 junctions produced by plasma- and by thermal- oxidation", Surface Science, Vol. 27, pp. 125-141, 1971) or rf sputtering in an oxygen plasma (see generally, J. H. Greiner, "Oxidation of lead films by rf sputter etching in an oxygen plasma", J. Appl. Phys., Vol. 45, No. 1, pp. 32-37, 1974). For the sake of brevity, some details noted above will not be repeated. The formation of the $Al/Al_2O_3/Al$ structures will

25 be seen to be simpler than that described for the Pb/PbO/Pb junctions owing to the much higher melting point of aluminum, relative to lead.

(i)   Sputter deposit aluminum on poly-Si at a temperature of ~25 to 150C. Due to thermodynamic forces, the micro-crystallites of the f.c.c. aluminum will have a strong and desirable (111) preferred orientation.

(ii)   Oxidize the aluminum *in situ* in molecular oxygen using temperatures, pressure and time to obtain the desired Al_2O_3 thickness. As with PbO, the thickness increases with log (time) and can be controlled via time at a fixed oxygen pressure and temperature to *within 0.10 Angstroms*, when averaged over a large number of aluminum grains that are present under the counter-electrode. One can readily change the Al_2O_3 thickness from ~15 to 35A by using appropriate oxidation parametrics (e.g., see Figure 2 in J. Grimblot and J. M. Eldridge, "II. Oxidation of Al films", J. Electro. Chem. Soc., Vol. 129, No. 10, pp. 2369-2372, 1982). The oxide will be amorphous and remain so until temperatures in excess of 400C are reached. The initiation of recrystallization and grain growth can be suppressed, if desired, via the addition of small amounts of glass forming elements (e.g., Si) without altering the growth kinetics or barrier heights significantly.

(iii) Re-evacuate the system and deposit a second layer of aluminum.

(iv)   Deposit the Poly-Si control gate layer using conventional processes.

Example III. Formation of Single- and Multi-Layer Transition Metal Oxide Tunnel Barriers.

Single layers of Ta_2O_5, TiO_2, ZrO_2, Nb_2O_5 and similar transition metal oxides can be formed by "low temperature oxidation" of numerous Transition Metal (e.g., TM oxides) films in molecular and plasma oxygen and also by rf sputtering in an oxygen plasma. The thermal oxidation kinetics of these metals have been studied for decades with numerous descriptions and references to be found in the book by Kubaschewski and Hopkins (O. Kubaschewski and B. E. Hopkins, "Oxidation of Metals and Alloys", Butterworth, London, pp. 53-64, 1962). In essence, such metals oxidize via logarithmic kinetics to reach thicknesses of a few to several tens

of angstroms in the range of 100 to 300C. Excellent oxide barriers for Josephson

tunnel devices can be formed by rf sputter etching these metals in an oxygen plasma

(see generally, J. M. Greiner, "Josephson tunneling barriers by rf sputter etching in

an oxygen plasma," J. Appl. Phys., Vol. 42, No. 12, pp. 5151-5155, 1971; O.

5   Michikami et al., "Method of fabrication of Josephson tunnel junctions," U.S. Pat.

4,412,902, Nov. 1, 1983). Such "low temperature oxidation" approaches differ

considerably from MOCVD processes used to produce these TM oxides. MOCVD

films require high temperature oxidation treatments to remove carbon impurities,

improve oxide stoichiometry and produce recrystallization. Such high temperature

10   treatments also cause unwanted interactions between the oxide and the underlying

silicon and thus have necessitated the introduction of interfacial barrier layers. See,

for example, H. F. Luan et al., "High quality Ta_2O_5 gate dielectrics with $T_{ox,eq}$ <10

angstroms," IEDM Tech. Digest, pp. 141-144, 1999.

A new approach was described in a copending application by J. M. Eldridge,

15   entitled "Thin Dielectric Films for DRAM Storage Capacitors," patent application

Serial No. 09/651,380 filed Aug. 29, 2000 that utilizes "low temperature oxidation"

to form duplex layers of TM oxides. Unlike MOCVD films, the oxides are very

pure and stoichiometric as formed. They do require at least a brief high temperature

(est. 700 to 800C but may be lower) treatment to transform their microstructures

20   from amorphous to crystalline and thus increase their dielectric constants to the

desired values (> 20 or so). Unlike MOCVD oxides, this treatment can be carried

out in an inert gas atmosphere, thus lessening the possibility of inadvertently

oxidizing the poly-Si floating gate. While this earlier disclosure was directed at

developing methods and procedures for producing high dielectric constant films for

25   storage cells for DRAMs, the same teachings can be applied to producing thinner

metal oxide tunnel films for the flash memory devices described in this disclosure.

The dielectric constants of these TM oxides are substantially greater (>25 to 30 or

more) than those of PbO and Al_2O_3. Duplex layers of these high dielectric constant

oxide films are easily fabricated with simple tools and also provide improvement in device yields and reliability. Each oxide layer will contain some level of defects but the probability that such defects will overlap is exceedingly small. Effects of such duplex layers were first reported by one J. M. Eldridge of the present authors and are

5   well known to practitioners of the art. It is worth mentioning that highly reproducible TM oxide tunnel barriers can be grown by rf sputtering in an oxygen ambient, as referenced above (see generally, J. M. Greiner, "Josephson tunneling barriers by rf sputter etching in an oxygen plasma," J. Appl. Phys., Vol. 42, No. 12, pp. 5151-5155, 1971; O. Michikami et al., "Method of fabrication of Josephson

10   tunnel junctions," U.S. Pat. 4,412,902, Nov. 1, 1983). Control over oxide thickness and other properties in these studies were all the more remarkable in view of the fact that the oxides were typically grown on thick (e.g., 5,000 A) metals such as Nb and Ta. In such metal-oxide systems, a range of layers and suboxides can also form, each having their own properties. In the present disclosure, control over the

15   properties of the various TM oxides will be even better since we employ very limited (perhaps 10 to 100 A or so) thicknesses of metal and thereby preclude the formation of significant quantities of unwanted, less controllable sub-oxide films. Thermodynamic forces will drive the oxide compositions to their most stable, fully oxidized state, e.g., Nb_2O_5, Ta_2O_5, etc. As noted above, it will still be necessary to

20   crystallize these duplex oxide layers. Such treatments can be done by RTP and will be shorter than those used on MOCVD and sputter-deposited oxides since the stoichiometry and purity of the "low temperature oxides" need not be adjusted at high temperature.

Fairly detailed descriptions for producing thicker duplex layers of TM oxides have been given in the copending application by J. M. Eldridge, entitled "Thin

25   Dielectric Films for DRAM Storage Capacitors," patent application Serial No. 09/651,380 filed Aug. 29, 2000, so there is no need to repeat them here. Although

perhaps obvious to those skilled in the art, one can sketch out a few useful fabrication guides:

(i)   Thinner TM layers will be used in this invention relative to those used to form DRAMs. Unlike DRAMs where leakage must be eliminated, the duplex oxides used here must be thin enough to carry very controlled levels of current flow when subjected to reasonable applied fields and times.

(ii)   The TM and their oxides are highly refractory and etchable (e.g., by RIE). Hence they are quite compatible with poly-Si control gate processes and other subsequent steps.

(iii)   TM silicide formation will not occur during the oxidation step. It could take place at a significant rate at the temperatures used to deposit the poly-Si control gate. If so, several solutions can be applied including:

(i)   Insert certain metals at the TM/poly-Si boundaries that will prevent inter-diffusion of the TM and the poly-Si.

(ii)   Completely oxide the TMs. The electrical characteristics of the resulting poly-Si/TM oxide 1/TM oxide 2/poly-Si structure will be different in the absence of having TM at the oxide/metal interfaces.

Example IV. Formation of Alternate Metal Compound Tunnel Barriers.

Although no applications may be immediately obvious, it is conceivable that one might want to form a stack of oxide films having quite different properties, for example, a stack comprised of a high dielectric constant (k) oxide/ a low k oxide/ a high k oxide. "Low temperature oxidation" can be used to form numerous variations of such structures. While most of this disclosure deals with the formation and use of stacks of oxide dielectrics, it is also possible to use "low temperature oxidation" to form other thin film dielectrics such as nitrides, oxynitrides, etc. that could provide additional functions such as being altered by monochromatic light, etc. These will not be discussed further here.

Example V. Formation of Perovskite Oxide Tunnel Barriers.

Some results have been obtained which demonstrate that at least a limited

range of high temperature, super-conducting oxide films can be made by thermally

5   oxidizing Y-Ba-Cu alloy films (see generally, Hase et al., "Method of

manufacturing an oxide superconducting film," U.S. Pat. 5,350,738, Sept. 27, 1994).

The present inventors have also disclosed how to employ "low temperature

oxidation" and short thermal treatments in an inert ambient at 700C in order to form

a range of perovskite oxide films from parent alloy films (see generally, J. M.

10   Eldridge, "Low Cost Processes for Producing High Quality Perovskite Dielectric

Films," application Serial No. _____). The dielectric constants of crystallized,

perovskite oxides can be very large, with values in the 100 to 1000 or more range.

The basic process is more complicated than that needed to oxidize layered films of

transition metals. (See Example III.) The TM layers would typically be pure metals

15   although they could be alloyed. The TMs are similar metallurgically as are their

oxides. In contrast, the parent alloy films that can be converted to a perovskite

oxide are typically comprised of metals having widely different chemical reactivities

with oxygen and other common gasses. In the Y-Ba-Cu system referenced above, Y

and Ba are among the most reactive of metals while the reactivity of Cu approaches

20   (albeit distantly) those of other noble metals. If the alloy is to be completely

oxidized, then thin film barriers such as Pd, Pt, etc. or their conductive oxides must

be added between the Si and the parent metal film to serve as: electrical contact

layers; diffusion barriers; and, oxidation stops. In such a case, the Schottky barrier

heights of various TM oxides and perovskite oxides in contact with various metals

25   will help in the design of the tunnel device. In the more likely event that the

perovskite parent alloy film will be only partially converted to oxide and then

covered with a second layer of the parent alloy (recall the structure of Figure 2),

then the barrier heights will represent that developed during oxide growth at the

parent perovskite alloy/perovskite oxide interface. Obviously, such barrier heights cannot be predicted *ab initio* for such a wide class of materials but will have to be developed as the need arises. This information will have to be developed on a system-by-system basis.

Methods of Operation

Write Operation

Write can be achieved by the normal channel hot electron injection and gate current through the silicon oxide to the floating gate. This is done by selecting a particular column by applying a high control gate voltage and applying relatively large drain voltage as is done with conventional ETOX flash memory devices. However, according to the teachings of the present invention, write can also be accomplished by applying a positive voltage to the substrate or well select line and a large negative voltage to the control gates, electrons will tunnel from the control gate to the floating gate. The low tunnel barrier will provide an easy write operation and the selection of the substrate or well bias will provide selectivity and address only one device.

Erase Operation

According to the teachings of the present invention, erase is achieved by providing a negative voltage to the substrate or well address line and a large positive voltage to the control gate. This causes electrons to tunnel off of the floating gate on to the control gate. A whole row can be erased by addressing all the column lines along that row and a block can be erased by addressing multiple row back gate or substrate/well address lines.

Read Operation

 Read is accomplished as in conventional ETOX flash memory devices. A column line is addressed by applying a positive control gate voltage and sensing the current along the data bit or drain row address line.

5

System Level

 Figure 8 shows a conventional NOR decode array 800 for memory circuits. The address lines are A1 through A3 and inverse address lines, $\overline{A1}$ through $\overline{A3}$. The conventional NOR decode array is programmable at the gate mask level by either fabricating a thin oxide gate transistor, e.g. transistors 801-1, 801-2, . . ., 801-N, at the intersection of lines in the array or not fabricating a thin oxide gate transistor, e.g. missing thin oxide transistors, 802-1, 802-2, . . ., 802-N, at such an intersection. As one of ordinary skill in the art will understand upon reading this disclosure, the same technique is conventionally used to form other types of decode arrays not shown. As shown in Figure 8, a number of depletion mode NMOS transistors, 816, are used as load devices.

 In this embodiment, each of the row lines 814 acts as a NOR gate for the address lines A1 through A3 and inverse address lines, $\overline{A1}$ through $\overline{A3}$ that are connected to the row lines 814 through the thin oxide gate transistor, e.g. transistors 801-1, 801-2, . . ., 801-N, of the array. That is, row line R1 is maintained at a high potential, +VDD, in the positive logic NMOS decode array shown in Figure 8A, unless one or more of the thin oxide gate transistor, e.g. transistors 801-1, 801-2, . . ., 801-N, that are coupled to row line R1 are turned on by a high logic level signal, +VDD, on one of the address lines A1 through A3 or inverse address lines, $\overline{A1}$ through $\overline{A3}$. When a transistor gate address is activated, by the high logic level signal, +VDD, through address lines A1 through A3 or inverse address lines, $\overline{A1}$ through $\overline{A3}$, each thin oxide gate transistor, e.g. transistors 801-1, 801-2, . . ., 801-N, conducts, or is turned "on." This conduction of the thin oxide gate transistor, e.g.

transistors 801-1, 801-2, . . ., 801-N, performs the NOR positive logic circuit

function, an inversion of the OR circuit function results from inversion of data onto

the row lines 814 through the thin oxide gate transistor, e.g. transistors 801-1, 801-2,

. . ., 801-N, of the array, in order to output a low logic level signal on the row lines

5   814. Thus, a particular row line 814 is addressed when none of the thin oxide gate

transistor, e.g. transistors 801-1, 801-2, . . ., 801-N, coupled to that row line 814 are

turned "on."

Again, the incoming address on each line is inverted and the combination of

the original address and inverted or complemented values used to drive the gates of

10   transistors in the decode array 800. The transistors 801-1, 801-2, . . ., 801-N in the

array 800 are enhancement mode NMOS devices and depletion mode NMOS

transistors are used as load devices 816. All voltages are positive in a simple

NMOS circuit. This is a positive logic NOR decode array, the logic one state, "1" is

the most positive voltage, +VDD, and the logic level zero, "0" is the least positive

15   voltage or ground.

The transistors used in Figure 8 are NMOS driver transistors with a

depletion mode NMOS load technology. The load device or NMOS load transistor

is a depletion mode or normally "on" transistor which acts as a constant current

source during the pull up switching transient thus providing high switching speed.

20   The driver transistor is an enhancement mode NMOS transistor which is normally

"off" with zero gate bias.

Figure 9 is a schematic diagram illustrating one embodiment of a decode

circuit, or memory address decoder, 900 according to the teachings of the present

invention. Analogous to Figure 8, the address lines are A1 through A3 and inverse

25   address lines, $\overline{A1}$ through $\overline{A3}$. As shown in Figure 9, the decode circuit 900 is

programmable at the gate mask level by either fabricating a driver transistor, or

logic cell, e.g. transistors 901-1, 901-2, . . ., 901-N, at the intersection of lines in the

array or not fabricating a driver transistor, or logic cell, e.g. missing floating gate

driver transistors 902-1, 902-2, . . ., 902-N, at such an intersection. In one

embodiment according to the teachings of the present invention, fabricating a driver

transistor, e.g. transistors 901-1, 901-2, . . ., 901-N, at the intersection of lines in the

array includes fabricating the floating gate driver transistor according to the

5   embodiments discussed and described in detail in connection with Figures 5 and 6.

In one embodiment of the present invention, as shown in Figure 9, a number of p-

channel metal oxide semiconductor (PMOS) load transistors, 916, are used as load

devices and are coupled to the output lines, or row lines, 914, of the decode circuit

900.

10   The incoming address on each address line A1 through A3 is inverted and

the combination of the original address on each address line A1 through A3 and

inverted or complemented values on inverse address lines, $\overline{A1}$ through $\overline{A3}$, used to

drive the control gates of transistors 901-1, 901-2, . . ., 901-N in the decode array

900. The floating gate driver transistors, or logic cells, e.g. transistors 901-1, 901-2,

15   . . ., 901-N in the decode array 900 are n-channel floating gate driver transistors

having a low tunnel barrier intergate insulator between the control gate and floating

gate as described above in connection with Figures 5 and 6.

In Figure 9, each of the row lines 914 acts as a NOR gate for the address

lines A1 through A3 and inverse address lines, $\overline{A1}$ through $\overline{A3}$ that are connected to

20   the row lines 914 through the floating gate driver transistors, or logic cells, e.g.

transistors 901-1, 901-2, . . ., 901-N, of the array 900. That is, row line R1 is

maintained at a high potential VDD, or logic "1" unless one or more of the floating

gate driver transistors, or logic cells, e.g. transistors 901-1, 901-2, . . ., 901-N, that

are coupled to row line R1 are turned on by a high logic level signal, VDD, on one

25   of the address lines A1 through A3 or inverse address lines, $\overline{A1}$ through $\overline{A3}$. In the

decode circuit 900 configuration shown in Figure 9, a logic "1", or VDD, on one of

the address lines A1 through A3 or inverse address lines, $\overline{A1}$ through $\overline{A3}$, is

required in order to turn on one of the n-channel floating gate driver transistors, or

logic cells, e.g. transistors 901-1, 901-2, . . ., 901-N, coupled to row line R1. As one

of ordinary skill in the art will understand upon reading this disclosure, the floating

gate driver transistors, or logic cells, e.g. transistors 901-1, 901-2, . . ., 901-N, can be

programmed to have two different conductivity states depending upon whether

5   electrons are stored on the floating gate. When a charge is stored on the floating

gate for any one of these floating gate driver transistors, 901-1, 901-2, . . ., 901-N,

the floating gate transistor is effectively removed from the programmable memory

address and decode circuit 900.

For the decode circuit 900 of the present invention, shown in Figure 9, the

10   driver transistors, e.g. transistors 901-1, 901-2, . . ., 901-N in the array are floating

gate transistor devices. In one embodiment, the floating gate driver transistors 901-

1, 901-2, . . ., 901-N are formed according to the embodiments of the present

invention as disclosed and described in detail in connection with Figures 5 and 6. In

this manner, the floating gate driver transistors, 901-1, 901-2, . . ., 901-N, can be

15   programmed initially in fabrication and can be reprogrammed as necessary once the

decode array is in service, e.g. field programmable, to implement a specific decode

function. The load devices 916, shown in the address decoder 900 of Figure 9, are

p-channel metal oxide semiconductor (PMOS) transistors and not depletion mode n-

channel transistors as is more usual. In this manner, the decode circuit 900

20   embodiment of the present invention shown in Figure 9 is formed according to a

CMOS process and can be referred to as a CMOS decode array 900.

In one embodiment, as shown in Figure 9, the decode circuit 900 of the

present invention includes at least one redundant row line, RD. In the embodiment

shown in Figure 9, a number of additional floating gate driver transistors, e.g.

25   transistors T1-T6, are provided in the array coupled to address lines A1 through A3

or inverse address lines, $\overline{A1}$ through $\overline{A3}$ and the redundant row line, RD. According

to the teachings of the present invention, these additional driver transistors, e.g.

transistors T1-T6, are formed according to the embodiments described and

discussed in detail above in connection with Figures 5 and 6. In one embodiment, as described above according to the teachings of the present invention, the additional floating gate driver transistors, T1-T6, will have a vertical control gate formed by the address lines A1 through A3 or inverse address lines, $\overline{A1}$ through $\overline{A3}$.

5   In another embodiment, as described above according to the teachings of the present invention, the additional floating gate driver transistors, T1-T6, will have a horizontal control gate formed by the address lines A1 through A3 or inverse address lines, $\overline{A1}$ through $\overline{A3}$ located above the floating gates of the floating gate driver transistors, T1-T6. According to the teachings of the present invention, the

10   second source/drain region for the additional driver transistors, T1-T6, are coupled to the at least one redundant row line, or wordline, RD. A p-channel metal oxide semiconductor (PMOS) load transistor T7, similar to p-channel metal oxide semiconductor (PMOS) load transistors 916 is coupled to the at least one redundant row line, RD as well to complete the CMOS inverter configuration.

15   As has been shown and described above, these non volatile, floating gate driver transistors, e.g. transistors T1-T6, can be programmed to have two different conductivity states depending upon whether electrons are stored on the vertical floating gate. When a charge is stored on the floating gate for any one of these floating gate driver transistors, e.g. transistors T1-T6, the floating gate transistor is

20   effectively removed from the programmable memory address and decode circuits 900 of the present invention. The implementation of these floating gate driver transistors, e.g. transistors T1-T6, in the decode circuit 900 of the present invention, enables error correction for replacing a row, or column in the array as one of ordinary skill in the art will understand upon reading this disclosure.

25   According to the teachings of the present invention, it is desirable to have redundant row lines, e.g. redundant row line RD, available to replace or error correct for row lines 914, which are determined defective or which have failed in the field.

The present invention is usable to provide such error correction by replacing a row, or column, in a memory decode circuit 900.

One of ordinary skill in the art will understand upon reading this disclosure that there can be more than one redundant row line, e.g. a RD2, RD3, etc. (not shown), and similarly more additional floating gate driver transistors, like transistors T1-T6, coupled thereto in order to enable multiple row error correction. One of ordinary skill in the art will further understand, upon reading this disclosure, the manner in which the additional floating gate driver transistors, T1-T6, formed according to the teachings of the present invention, can be selectively programmed in order to access, or select, redundant rows RD in replacement for any one of the output lines 914 in the decode array 900.

In summary, If electrons are stored on a floating gate for one of the additional floating gate driver transistors, T1-T6, then when a high input signal is received on address lines A1 through A3 or inverse address lines, $\overline{A1}$ through $\overline{A3}$, the "programmed floating gate driver transistor, T1-T6, will remain "off." On the other hand, if there is no stored charge on the vertical floating gate for that particular floating gate driver transistors, T1-T6, then the floating gate driver transistors, T1-T6, will conduct when a high input signal is received on address lines A1 through A3 or inverse address lines, $\overline{A1}$ through $\overline{A3}$ associated with that floating gate driver transistor. If the floating gate driver transistors, T1-T6, have no charge stored on the floating gate they will function as normal inverters for the decode circuit 900. Conversely, if there is a charge stored charge on the floating gate, the conductivity of the floating gate driver transistors, T1-T6, will not become high enough and will not function as a driver transistor. In this latter case, the output for the redundant row line RD in the decode circuit 900 of the present invention will not change charge states. Hence, if there is a charge stored on the vertical floating gate of the floating gate driver transistors, T1-T6, the drivers are effectively removed from the decode circuits 900.

Analogously, the decode circuit shown in Figure 9 can represent a column decode circuit 900. In this case, the lines 914 or redundant line RD which are coupled to the address lines A1 through A3 or inverse address lines, $\overline{A1}$ through $\overline{A3}$ through the floating gate driver transistors, 901-1, 901-2, . . ., 901-N and T1-T6, can

5   be complementary bit lines for column decoding as the same will be know and understood by one of ordinary skill in the art.

As one of ordinary skill in the art will further understand upon reading this disclosure, additional inverters can be used as necessary to affect the transition from one logic system, e.g. positive logic system, to a negative logic system while still

10   capitalizing on the utility of the novel floating gate driver transistors 901-1, 901-2, . . ., 901-N and T1-T6, having a low tunnel barrier intergate insulator between the control gate and floating gate in decode circuit 900. If the floating gate in a floating gate driver transistor is programmed with a negative charge on the floating gate it will not be active in the array and it is effectively removed from the array. In this

15   manner the array logic functions can be programmed even when the circuit is in the final circuit or in the field and being used in a system. The field programmable, in service or in circuit programmable, logic devices described here work with much lower voltages than the normal devices used in current in field, or in service, programmable decode circuit technology due to tunneling of charge between the

20   floating gate and control gate through the low-tunnel barrier intergate insulator. They can be programmed with Voltages of 2.0 to 4.0 Volts and the normal operating voltages on the vertical control gates can be of the order 1.0 Volt or so.

The absence of presence of stored charge on the floating gates is read by addressing the x-address or control gate lines and y-column/sourcelines to form a

25   coincidence in address at a particular floating gate. The control gate line would for instance be driven positive at some voltage of 1.0 Volts and the y-column/sourceline grounded, if the floating gate is not charged with electrons then the floating transistor would turn on tending to hold the row or word address line on that

particular row down indicating the presence of a stored "one" in the cell. If this particular floating gate is charged with stored electrons, the transistor will not turn on and the presence of a stored "zero" indicated in the cell. In this manner, data stored on a particular floating gate can be read. In reality, data is read out in "bit pairs" by addressing not only a single floating gate but rather both of the floating gates in row adjacent pillars on each side of a particular control gate address line. Data is stored into the cell by hot electron injection or alternatively, according to the teachings of the present invention, by tunneling charge from the control gate to the floating gate through the low-tunnel barrier intergate insulator. Erasure is accomplished by driving the control gate line with a positive voltage and the sourceline, or backgate/substrate well address line, of the transistor with a negative bias so the total voltage difference is in the order of 2 Volts causing electrons to tunnel off of the floating gates and to the control gates. According to the teachings of the present invention, data can be erased in "bit pairs" since both floating gates on each side of a control gate can be erased at the same time. This architecture is amenable to block address schemes where sections of the array are erased and reset at the same time.

Figure 10 is a simplified block diagram of a high-level organization of an electronic system 1001 according to the teachings of the present invention. As shown in Figure 10, the electronic system 1001 is a system whose functional elements consist of an arithmetic/logic unit (ALU) 1020 or processor 1020, a control unit 1030, a memory device unit 1040 and an input/output (I/O) device 1050. Generally such an electronic system 1001 will have a native set of instructions that specify operations to be performed on data by the ALU 1020 and other interactions between the ALU 1020, the memory device unit 1040 and the I/O devices 1050. The memory device units 1040 contain the data plus a stored list of instructions.

The control unit 1030 coordinates all operations of the processor 1020, the memory device 1040 and the I/O devices 1050 by continuously cycling through a

set of operations that cause instructions to be fetched from the memory device 1040 and executed. Memory device 1040 can be implemented with "in-service" programmable low voltage decode circuits, according to the teachings of the present invention, e.g., having a low-tunnel barrier intergate insulator between the control

5   gate and floating gate. In addition, the decode circuits of the present invention can enable error correction by replacing a row, or column, in a memory array.

CONCLUSION

The above structures and fabrication methods have been described, by way

10   of example and not by way of limitation, with respect to in service programmable logic arrays using non-volatile memory cells with low tunnel barrier interpoly insulators.

It has been shown that the low tunnel barrier interpoly insulators of the present invention avoid the large barriers to electron tunneling or hot electron

15   injection presented by the silicon oxide-silicon interface, 3.2 eV, which result in slow write and erase speeds even at very high electric fields. The present invention also avoids the combination of very high electric fields and damage by hot electron collisions in the which oxide result in a number of operational problems like soft erase error, reliability problems of premature oxide breakdown and a limited number

20   of cycles of write and erase. Further, the low tunnel barrier interpoly dielectric insulator erase approach, of the present invention remedies the above mentioned problems of having a rough top surface on the polysilicon floating gate which results in, poor quality interpoly oxides, sharp points, localized high electric fields, premature breakdown and reliability problems.

25   According to the teachings of the present invention, any arbitrary combinational logic function can be realized in the so-called sum-of-products form. A sum of products may be implemented by using a two level logic configuration such as the NOR-NOR arrays shown in Figure 10, or by a combination of NOR

gates and NAND gates. A NAND gate can be realized by a NOR gate with the inputs inverted. By programming the floating gates of the non-volatile memory cells in the array, these arrays can be field programmed or erased and re-programmed to accomplish the required logic functions.

5